USABILITY OF NON-SPEECH SOUNDS IN USER INTERFACES

Rafa Absar

CIRMMT and McGill
SIS, 3459 McTavish St.
Montreal, QC H3A 1Y1, Canada
rafa.absar@mail.mcgill.ca

Catherine Guastavino

CIRMMT and McGill
SIS, 3459 McTavish St.
Montreal, QC H3A 1Y1, Canada
catherine.guastavino@mcgill.ca

ABSTRACT

We review the literature on the integration of non-speech sounds to visual interfaces and applications from a usability perspective and subsequently recommend which auditory feedback types serve to enhance human interaction with computers by conveying useful and comprehensible information. We present an overview over varied tasks, functions and environments with a view to establishing the best practices for introducing non-speech sounds in order to improve the overall experience of users.[1]

1. INTRODUCTION

Humans use a combination of vision and audition in our everyday lives to gather information from our surroundings and interact with the world. Such a combination can be a powerful tool for interaction with human-computer interfaces. However, common interfaces today are mostly graphically-oriented; very little information is presented via the other modalities [1]. Hence, the myriad advantages that can be offered by integrating visual and audio cues are not yet fully realized.

While vision provides detailed information, it requires direct focus on the area providing information, whereas audition can provide more general information even outside peripheral vision [1]. If all the information is presented visually, it may lead to visual overload [2] and may also lead to some information being missed, if the eyes are focused elsewhere. Brown *et al* [2] suggested dividing this information by replacing some of the cues traditionally presented in the visual modality with auditory cues, so as to reduce some of the visual workload. It has been found that auditory cues indicating the location of objects improve visual search speed and accuracy, for example, in tasks where a visual target has to be located on the screen [2], [3].

Some of the advantages offered by nonspeech sounds besides reducing workload on users' visual system [2] are that they can provide complementary information to vision [1]; they can also reduce the amount of information needed to be displayed on screen, hence optimizing screen space and reduce the demand on visual attention. Sound is also attention-grabbing and can be used for peripheral awareness and ambient audio [4].

However, there are some disadvantages to using nonspeech sounds [1] which need to be addressed if audio is to be successfully incorporated into human-computer interfaces. Presenting either abstract data or absolute data using sound is often difficult. Sounds can be used to portray relative differences in values, but to get an absolute value, users typically require looking at a number or graph. Another issue is that audio is a transient medium which disappears after it is presented, and has to be replayed if not remembered. Stationary visual data, on the other hand, can be referred back to whenever required. Many audio parameters are unsuitable for high-resolution display of information. And finally, auditory feedback can cause annoyance in users if not designed appropriately.

The field of auditory display is a relatively novel discipline and therefore could benefit from a comprehensive analysis of literature and best practices in the area of auditory feedback for visual interfaces. Hence, the analysis provided as part of this paper is intended to provide a valuable resource for the auditory display community.

Nonspeech sound used for auditory feedback can be divided into two main categories. Sounds that can be easily attributed to objects or events generating sounds in everyday situations are referred to as auditory icons [5]. Abstract sounds, typically synthetic and less identifiable, are referred to as earcons [6]. Both types of sounds are described in the next two subsections. Following this, studies addressing audio-visual integration and evaluations of the two types of auditory feedback and how they affect task performance or user experience, are discussed (Section 2). Section 3 delves into descriptions of several tools and applications that have used non-speech auditory feedback. Section 4 summarizes and concludes the discussion.

1.1. Auditory icons

The concept of auditory icons was introduced by Bill Gaver [5] as emulations or caricatures of naturally occurring sounds in everyday life. Gaver suggests that humans perceive everyday sounds in terms of the sources, materials and actions that made them, rather than the individual sound attributes such as pitch and timbre [7]. Hence, an auditory icon is a sound that is intended to provide information about an event or object in the interface by representing the desired data using properties of the sound's source, rather than properties of the sound itself [5]. Another important property of such everyday sounds is that they convey information about the sound source (e.g. size and material), and the interaction (e.g. force applied or action). These features can be useful in providing multi-faceted information in human-computer interfaces [1]. For a review of the perception of everyday environmental sounds, see [8].

One of the earliest applications using auditory icons is the SonicFinder [9]. Gaver used nonspeech real-life sounds as auditory feedback to interface events that can be intuitively be mapped to the respective sounds as analogies to the actions performed (di-

[1]This research was supported by FQRNT grant 113581 to C. Guastavino

rectly or metaphorically). For instance, selecting a file was mapped to the sound of an object being tapped, with the type of object indicated by the object material, and size of file represented by the size of the struck object. Since the design of auditory icons requires intuitive mapping to the computer interface model, this led Fernstrom *et al* to explore what people hear to develop an understanding of people's perception of auditory events so as to identify mappings and metaphors for actions in the interface [10], [11]. Most desktop interfaces today do implement certain forms of auditory icons as additional feedback to visual events, such as the metallic crunching sound that accompanies the action of placing an object in trash (deletion). However, many of these implementations have not been formally evaluated.

1.2. Earcons

Earcons are abstract audio messages in the user-computer interface that provide feedback to the user [6]. The sounds used are synthetic combinations that are manipulated in structured ways by complex manipulations of the parameters of sound, such as the timber, pitch and rhythm. This process allows the representations of hierarchies to create both simple and complex audio messages to encode information at the interface level.

While representational auditory icons may have the advantage of being easier to remember and learn as they sound more familiar and relatable, abstract earcons have a number of advantages working for them as well: systematic, well-defined building blocks can be used to create larger sets of earcons more easily [6]; families and hierarchies can be created out of basic audio messages, unlike auditory icons. It has been shown that such structured audio messages can reduce information overload by improving usability and task performance, e.g. by reducing the time to recover from errors [12].

1.3. Summary

Nonspeech audio provides a valuable medium to communicate information and provide feedback to the user, although its shortcomings have to be addressed to be successfully integrated in computer interfaces. There are two types of auditory cues: auditory icons are emulations of real-life sounds portraying different functions and events in the interface; earcons are abstract, synthetic sounds. The advantages and disadvantages of each type of sounds will be discussed in more detail in the following sections.

2. EVALUATIONS OF NON-SPEECH SOUNDS

There has been extensive research in the domain of psychological and perceptual studies investigating the integration of auditory and visual signals [13], [14], [15], [16], [17]. Most of these studies have employed methods that examine the interaction of audio and vision at a low level of processing, using simplified synthetic stimuli such as light flashes, beeps, noise bursts or pure tones. Although it is difficult to link the results of these studies directly to the more complex processing of audio-visual information at the interface level, most support the notion of audio-visual integration in computer interfaces. Presenting bimodal information has been found to enhance perception or improve performance, as long as the information presented is relevant or complementary.

Research on selective attention in audio-visual bimodal presentation has investigated the processes by which people focus on information relevant to the task at hand and ignore what is irrelevant or distracting, in the two modalities. Studies have shown that this selective attention is brought about by the brain selectively increasing the sensitivity of the areas responsive to the task-relevant features and decreasing the sensitivity in the corresponding area responding to the non-relevant feature, thus serving to selectively enhance perception [13]. Bimodal presentation of complementary information often leads to better performance in certain perceptual tasks, such as object recognition as shown in [14]. In this study, Molholm *et al* [14] examined the combined influence of visual and auditory inputs on identification of objects - in this case, pictures and animal vocalizations. It was found that participants identified objects significantly faster and more accurately when the picture and vocalization matched, compared to unimodal presentations. However, Johnson and Zatorre [15] showed that when bimodal presentation was not complementary or when one modality did not provide useful information, no such performance improvement can be found.

Driver and Spence [16] used left-right discrimination of an auditory or visual target accompanied by a cue in another modality and observed that the task was facilitated when the relevant stimuli were presented from the same spatial location across the two modalities. Furthermore, Spence and Ranson found that it is more difficult to ignore distracting sounds when they are presented at a visually relevant location [17]. These findings provide guidelines for the effective design of multimodal interfaces by illustrating the potential trade-off between arrangements that make it easier to attend to simultaneous relevant information in multiple modalities in the same location, or conversely, more difficult to ignore irrelevant information. Hence, it may be beneficial to design multimodal interfaces in such a way that sounds that are not immediately relevant to a current task are spatially located away from the area of visual focus. This means spatially distributing sounds that are not immediately relevant to the task at hand in an interface in order to minimize the distracting effect.

2.1. Methods

The methods used in experiments to evaluate the nonspeech auditory feedback are discussed by enumerating the following aspects of the experiments: the types of tasks that were designed, the independent variables and the dependent variables, including the types of subjective ratings.

Types of tasks designed for evaluations included:

- Finding and selecting the lowest level menu items or recalling menu levels in a hierarchical menu structure, guided by earcons, to evaluate if earcons are effective in communicating menu structures [18], [19], [20], [21].
- Navigating and localizing in a room-based simulation guided by nonspeech audio, having to follow a predetermined route and making sure everything is working properly using the background sounds, to evaluate if nonspeech audio helps in navigational support systems [22].
- Identifying picture categories made of line drawings by classifying the pictures as animals or non-animals, to investigate how task performance is affected by auditory icons and earcons. Some conditions employ a dual-task of including a mental addition task for greater cognitive load [23], [24], [25], [26], [27].
- Monitoring the simulation of a factory using background sounds [28].

- Listening to and describing everyday sounds [10], [29], or listening to and selecting the audio cues that best match a function or computer event [30], [31].
- Braking when an accident seems imminent in a vehicle-collision avoidance system using nonspeech audio cues and warnings [32].

Table 1 summarizes and outlines these methods, tasks and the independent and dependent variables involved; the references are cited in the above task list. The results are discussed in the following sections.

Table 1: Outline of experimental tasks and variables

Tasks	**Independent variables**	**Dependent variables**
Finding and selecting the lowest level menu items or recalling menu levels using earcons	Type of menu, icons, commands; Presence or absence of sound; Earcons as musical timbres, simple tones or sounds with no rhythm	No. of correct menu level identifications; error numbers and rates; ease of use
Navigating in a room-based simulation using the background sounds	Feedback as auditory icons, earcons, or no sound	No. of correct object recognitions; ease of use; annoyance factor
Identifying picture categories of animals or non-animals	Presence of auditory icons, earcons, no sound; relevance of cues	Response time
Monitoring the simulation of a factory using background sounds	Presence or absence of auditory icons; Task difficulty	Time for completion; no. of correct object identifications; recall performance; no. of errors
Listening to and describing everyday sounds, or mapping audio cues to a function		Recall performance; confidence in mapping; pleasantness; appropriateness
Braking when an accident seems imminent in a vehicle-collision avoidance system	Feedback as auditory icons or earcons	Response time; error rate; preference

2.2. Evaluating auditory icons

Gaver developed the ARKola simulation of a soft drinks bottling factory to evaluate the effectiveness of auditory icons. This type of auditory feedback was used to represent some of the objects and events within the simulation interface [28]. Pairs of users controlled the factory to test if the audio feedback allowed more efficiency in running the plant and if it affected collaboration efforts. It was observed that the feedback formed a combination of several sounds intermixed with one another, much like the ecology of sounds we use in everyday life, and led the users to be more effective in monitoring the status of ongoing processes. Users were not as efficient in the purely visual condition. The addition of auditory icon feedback also better allowed collaboration between users since one user could not always see the other user's part of the plant, but could hear and identify the relevant feedback.

Mynatt [29] investigated how well people identify auditory cues by asking participants to describe a collection of short everyday sounds. Results showed that people identified a sound only 15% of the time (a sound is said to be identified if it is mapped to the recorded action or source) in the absence of context. She also found that certain sounds were systematically identified as objects (such as cameras, printers, doors) and some as actions producing a sound (such as closing, tearing, locking). This supports the guideline that some sounds should be selected to represent interface objects, and others for actions.

Similar kinds of listening tests and free-text identification were carried out by Fernstrom *et al* ([10] and [11]). These responses and categorizations were gathered to investigate how accurately people can identify sounds, and suggest possible mappings and metaphors to the human-computer interface. They found that a fairly high percentage of objects and actions were correctly identified (about 70% of the time). Similar to [29], they also found that hearing sounds without context can be confusing, and the order of presentation of the sounds affects the way it was identified. Hence, auditory feedback has to be designed relevant to the function, with care being taken not to have loss of context or high ambiguity.

Graham [32] on the other hand, evaluated the use of auditory icons for in-vehicle collision avoidance applications. He compared auditory icon warnings with earcons and speech warnings as well. He measured the braking reaction times, number of inappropriate responses and subjective ratings of participants. Results showed that although auditory icon warnings gave faster reaction times and were also rated higher subjectively, they did result in a higher number of inappropriate responses. This meant that the perceived urgency and inherent meaning of such everyday sounds can be easily misinterpreted and care needs to be taken to design these sounds as warnings for such critical applications.

2.3. Evaluating earcons

Earcons differ from auditory icons in that they have no natural link or mapping to the objects or events they represent, and hence, have to be learned.

Brewster *et al* performed a series of detailed experiments based on compound and hierarchical earcons to examine their effectiveness [33], [34]. Participants were presented with earcons representing families of icons, menus or both, and had to identify them when played back. This study also investigated whether musical ability affected recall performance and it was found that earcons were recalled equally well by both musicians and non-musicians. However, training the participants in familiarizing themselves with the sounds used was an important factor in recall performance. They conclusively found that earcons were more effective than unstructured bursts of sound and that musical sounds were more effective than simple tones, which differed significantly from the design principles proposed by Blattner *et al* [6]. A richer design based on more complex musical timbres gave even better results in communicating information in computer interfaces, leading to the conclusion that complex sounds should be used to design earcons rather than simple tones [33].

Barfield *et al* [18] studied whether earcons can help to represent and recall depth in a menu structure and found that sound did not improve recall performance of depth in menu structures and users found them distracting. However, in a later study, Brewster *et al* [19] also investigated whether earcons effectively provide navigation cues in a menu hierarchy and found different results. Earcons were created for a hierarchy of menu levels and nodes, and participants had to identify their location in the hierarchy using these earcons. Results showed over 80% accuracy, providing evidence that earcons afforded an efficient method for menu localization cues. This study was further extended in [21] with a larger hierarchy, more types of earcons and with a test of recall of earcons over time. Recall over time had good results, but they did, however, find that the type of training had significant results on recall performance. In [21], it was found that lower sound quality lowered recall of earcons - in this case, CD quality sound over the lower quality of sound played over the telephone in telephone-based interfaces.

Lemmens *et al* [26] studied the effect of earcons in picture categorization tasks of animal and non-animal line drawings with auditory cues containing redundant information. The drawings were presented either with relevant information via auditory cues of sounds of animals or objects matching the picture, or non-relevant cues in the incongruent condition. In one of the experiments, participants had to carry out an additional mental addition task for greater cognitive load. Results showed earcons containing relevant redundant information helped reduce errors in both the single and dual-task environments.

In a similar study including picture categorization tasks [23], mood cues in major and minor chords were used with the pictures to see if they affected performance. It was hypothesized that earcons in minor chords suggest a negative emotion and hence should favour a negative answer, whereas those in major chords should favour a positive answer. According to this hypothesis, when the answer to the picture categorization task was positive (yes for animal), major chords should help in response time. However, the auditory mood cues seemed to delay responses in these tasks, leading to the conclusion that the auditory modality together with the visual modality was not always appropriate for these tasks, which they referred to as the modality appropriateness hypothesis. It is doubtful, however, whether this conclusion can be drawn in all similar situations, since the validity of the use of the auditory mood cues in these tasks is open to interpretation.

2.4. Comparisons and combinations

Both auditory icons and earcons have been found to be effective in communicating information in the human-computer interface through audition. However, each method has its own advantages and disadvantages, and no single method has been conclusively shown to be superior to the other.

Lucas [30] evaluated the two types of nonspeech auditory feedback, and compared them with speech cues as well. Participants had to listen to and select which audio cues from the three best represented an action or object in the interface. An explanation of the design of the cues were given to half the subjects and both halves tested again one week later to see if design knowledge helps recall performance. It was found that this prior knowledge did help retain information on the cues. Results also showed that after speech, auditory icons were most accurately associated with the correct action or object.

Bussemakers *et al* [24] investigated whether redundant auditory icons used with visual information influence the performance on picture categorization tasks on line drawings, and they compared the results with experiments using redundant earcons. Results illustrated that response times are faster in conditions with auditory icons than the silent condition, and that response times were even slower with earcons than the silent condition. Lemmens *et al* [27] performed two more similar experiments, one with a dual-task requiring a mental addition task with the picture categorization, and one experiment using intermixed auditory icons and earcons with the picture categorization task. These experiments confirmed the previous results: although the dual-task slowed reaction times, auditory icons still led to faster response times than earcons. Hence, auditory icons seemed to have a facilitatory effect in picture categorization tasks of this kind, while earcons seemed to have an inhibitory effect.

A navigational support approach in a building maintenance system using a room-based metaphor was evaluated in [22]. It used auditory icons for audio feedback and was compared with the system using earcons and no sound. None of the subjects preferred the earcon condition in this experiment, while some preferred auditory icons. Also, auditory icons allowed better recall performance.

Sikora *et al* [31] designed auditory feedback in a graphical user-interface for business communication using either musical sounds (earcons) or real world sounds (auditory icons). Users mapped the sounds to functions and rated their confidence in the functional mapping, its pleasantness and appropriateness. Real world sounds mapped most predictably to functions, although musical sounds had higher ratings for pleasantness. For the business application, no auditory icons were selected. Hence, preference does not always reflect the best functional mapping. The authors also concluded that real world sounds may be less appropriate for actual workplace applications. Edworthy [35] tried to determine if sound helps people work better with machines and suggested that real world sounds may be more suited for auditory feedback on monitoring tasks via background sounds, while abstract sounds may be better suited for warnings and alarms as they tend to attract our attention more effectively.

2.5. Summary

Table 2 summarizes the more relevant or definitive findings of the evaluative studies discussed in this section. Factory-monitoring tasks were facilitated by adding auditory icons; recall of level and location in hierarchical menus were improved by adding earcons. However in picture categorization tasks and room-based navigation tasks, auditory icons were found to be more effective and preferred than earcons. This is in contrast with business applications, where earcons were given higher subjective ratings, and in-vehicle collision avoidance systems, where earcons gave rise to fewer errors than auditory icons.

3. CONTEXT OF USE

In this section, some of the applications that have used nonspeech sounds for enhancement and feedback are discussed. Examples also include a few non-visual and audio-haptic interfaces as well.

Table 2: Summary of evaluative and comparative studies

Tasks or Environment	Type of sound	Results
Factory monitoring systems	Auditory icons	Increased efficiency, increased collaboration
In-vehicle collision avoidance system	Auditory icons	Low responses time, high number of inappropriate responses, high subjective ratings
	Earcons	Less number of inappropriate responses
Hierarchical menus	Earcons	Highly effective in recall of menu level and location
Picture categorization tasks (line drawings)	Auditory icons	Facilitatory effect
	Earcons	Inhibitory effect
Navigational support system in room-based simulations	Auditory icons	Higher subjective rating, higher recall
	Earcons	Lower subjective ratings and recall
Business applications	Auditory icons	Better functional mapping, low subjective rating
	Earcons	High subjective ratings

3.1. Desktop applications

One of the first desktop interfaces developed using auditory icons was the SonicFinder [9], mentioned in Section 1.1, where real-life sounds were mapped to different common interface objects and events for intuitive auditory feedback. For example, selecting interface objects made sounds of tapping a material depending on the type of object, e.g. files gave a wooden sound, applications a metal sound and folders a paper sound. Copying actions were aurally illustrated using a pouring analogy - the sound of how full the receptacle was indicated the progress of the copy action (with increasing pitch). The challenge, however, is finding representative sounds for all actions and events, since some events at the interface level are abstract and difficult to portray with a real-life sound.

Brewster *et al* [36] developed earcons for desktop use and performed detailed evaluations on different types of earcons, as discussed in Section 2.

3.2. Complex systems

Some applications that have utilized nonspeech audio have integrated it into much more complex environments than the desktop interface. One such study mentioned previously in Section 2.2 was the ARKola simulation, which used an ecology of auditory icons in a complex soft drinks factory simulation to convey information about the current state of the factory and its components and help improve collaboration efforts [28].

Skantze and Dahlback [22] described another such complex environment portraying a navigation support approach based on auditory icons for navigating in room-based designs. The prototype system simulated a buildings maintenance support system using a room-based metaphor. It was found that users responded positively to the use of auditory icons, rather than earcons, in this environment.

Mynatt *et al* [37] designed a more complex system that provides continuous serendipitous information to users via background auditory icon cues in the workplace. The Audio Aura system provides information to the user even when away from his desk, so that users do not have to be confined to their office space at all times. The auditory peripheral cues are meant to be ambient and provide information which can be ignored if not required. For example, the sound of surf represented the amount of new e-mails received by the user, with a higher number of e-mails being characterized by increasing surf. They used an electronic tag and networking system and wireless headphones linked to each person in the workplace for tracking and notifying purposes.

3.3. Mobile devices

In today's world where communication in a mobile environment is critical, mobile devices interfaces have to be designed well to compensate for the lack of screen space and low-resolution visual data. Hence, in [38] Brewster *et al* proposes the use of nonspeech sounds to improve interaction without the need for more screen space. Later, Leplatre and Brewster [39] describe a framework for integrating such nonspeech audio to mobile phone menus where visual feedback is constrained. The hierarchical menu structures were enhanced using earcons, and evaluations showed significant performance benefits from the sonifications.

In [40], a prototype audio user interface for a GPS system is designed so that users can carry out location tasks on mobile computers while their attention and hands are occupied elsewhere. The interface uses a simple form of spatial audio, rather than speech audio, and was shown to be effective and inexpensive for location tasks.

A similar application is the Nomadic radio, a wearable computing platform for managing voice and text-based messages in a mobile environment [41]. It uses an auditory user-interface for navigational and notification purposes among messages. Speech audio and spatial auditory icon cues are continuously played in the background to provide peripheral awareness of the system status. Evaluations showed that users preferred this type of auditory awareness to speech-based navigation systems.

3.4. Applications for the visually impaired

One of the most important uses, and the most widely studied areas, for nonspeech audio is in computer applications for visually-impaired users (for a recent review, see [42]). Since speech audio takes time to be played out and listened to, and hence is not the most efficient method of communication, nonspeech audio can be effectively replaced as some of the feedback in such applications. Mynatt [43] developed a methodology for transforming graphical interfaces into nonvisual auditory interfaces by converting the salient components of the graphical interfaces into auditory components. Auditory icons are used to convey these interface objects, based on a hierarchical model of a graphical interface, providing visually-impaired users many of the benefits of graphical user interfaces (GUIs). Mynatt and Weber [44] also compared two different applications for converting GUIs to nonvisual interfaces: *Mercator* replaces the spatial graph display with a hierarchical auditory interface, while *GUIB* translates the screen contents into tactile information based on the spatial arrangement of the GUI. User evaluations showed that auditory cues as used in Mercator were very effective for nonvisual interfaces.

Morley *et al* [45] designed an auditory system for visually-impaired users to enable efficient navigation on the web or hypermedia. This interface uses nonspeech sounds to identify links and provide information and feedback about text and commands to improve usability. They incorporated naturalistic auditory icons where appropriate, to engage blind students, and simple earcons for other situations. Evaluations showed that participants liked these sounds and found them easy to remember. The auditory feedback allowed them to work faster and more efficiently than conditions without feedback. Goose and Moller [46] also designed a web browser using spatialized 3D audio to convey the structure of the hypermedia document. It provides audio structural surveys, positional audio feedback of links and anchors, progress indicators and meta-information of new links, improving browsing experience for both sighted and visually-impaired users.

Another such tool for web access was developed by Murphy *et al* [47]. They designed a plugin for web browsers that provides auditory feedback and haptic cues to enable visually-impaired users to spatially localize themselves on web pages, and build a mental model of the spatial structure of the web document. The plugin generates audio feedback to indicate links, images, and other such web objects, and also aurally indicates when the user crosses the boundaries of the page.

3.5. Immersive systems

Auditory feedback also plays a very important role in contributing to the feeling of presence in immersive virtual environments [48]. Grohn *et al* [49] carried out a navigation test in a spatially immersive virtual environment that simulated a game-like experience to test this. The system used both auditory and visual cues. It was found that audio-visual navigation was more efficient and immersive, than only visual or auditory navigation in a 3D virtual environment.

Auditory feedback has been added to virtual assembly environments and studies have been performed to evaluate task performance in such environments. Zhang and Sotudeh [50] presented an approach for the integration of 3D auditory feedback into virtual assembly environments and evaluated the resulting system. They reported that the addition of auditory feedback improved task performance and that audio-visual integration gave the best results, when compared to any individual modality feedback alone. Edwards *et al* [51], on the other hand, studied whether the inclusion of auditory cues or force feedback to an immersive virtual environment improved the performance of an assembly task. Results showed that the addition of force feedback slowed completion times and increased errors in some users, while auditory feedback had no such negative performance effects.

3.6. Summary

Applications and tools that apply auditory feedback to improve usability of desktop applications, web interfaces, or more complex environments simulating real-life situations are discussed. The applicability of auditory feedback to mobile devices is discussed to improve usability by reducing visual clutter and amount of screen space required to communicate information. Applications using audio feedback for better access by visually-impaired users are also described. Applying auditory enhancements to immersive virtual environments has also been found to increase the sense of presence and improve performance in virtual assembly tasks as well as navigation tasks.

4. CONCLUSION

Previous research shows that nonspeech audio is an effective means of communicating information to the user in the computer interface, be it via auditory icons or earcons, in a multitude of applications. The results of the studies discussed here are promising for audio-visual integration in computer interfaces, as relevant auditory feedback tends to enhance task performance in the specific modalities.

Table 3: Applications and tasks where the addition of auditory icons have been found to facilitate tasks

Desktop interfaces	Navigation; picture categorization; hypermedia and web interfaces
Complex systems	Monitoring tasks; collaborative tasks; peripheral awareness cues, ambient sound; navigation tasks
Immersive virtual environments	Localization and navigation tasks; assembly tasks

Table 4: Applications and tasks where the addition of earcons have been found to facilitate tasks

Desktop interfaces	Sonically-enhanced widgets; menu hierarchies; business and workplace applications; graphs and tables; hypermedia and web interfaces
Alarms and warning systems	Vehicle-collision detection
Immersive virtual environments	Assembly tasks
Mobile systems	Mobile phone menus

Auditory icons have the advantage of being easy to learn and remember as they are natural and relatable to our everyday lives. Audio messages with good mappings and metaphors can make for a very effective feedback system for most users. However, the disadvantage for this type of feedback also arises from this issue. All computer interface functions and objects do not have real world equivalents and it may be hard to find a metaphor to represent such functions without being faced with issues of ambiguity, loss of context and even annoyance factors in users.

While earcons have the converse disadvantage or having to be learned and remembered since they have no natural intuitive link to the interface action or object, they have the advantage of being highly structured. As such, it is easier to follow structured design principles to create families of earcons, so that users typically can learn to recognize them by remembering their common characteristics and attributes. Auditory icons, on the other hand, have to be remembered individually, as it is not easy to connect them in structured families.

This paper has highlighted the fact that auditory icons and earcons are each more effective than the other in different environments and task situations. Preliminary conclusions and deductions that can be drawn from the literature reviewed has been categorized and summarized in Table 3 and Table 4, for auditory icons and earcons, respectively. However, there has been no entirely conclusive evidence to be able to say that one method of feedback is most certainly better than the other in a certain environment. Most have only been found to be effective in very specialized and specific applications or task situations. Hence, for future research, it would be worth investigating these functional mappings more closely, design more general tasks and environments, and evaluate if they are truly optimal for those respective contexts.

There has been more formal research into the design of earcons and the evaluation of systems using them than there has been into auditory icons. Hence, designing formal evaluations of more functions and utilities of auditory icons in different task environments, in the context of human-computer interfaces, also merits further research. Further research is required to combine the two types of auditory feedback in a single system, by using both types of sounds to their full potential and capabilities.

5. REFERENCES

[1] S. Brewster, "Nonspeech auditory output," *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, pp. 220–239, 2003.

[2] M. L. Brown, S. L. Newsome, and E. P. Glinert, "An experiment into the use of auditory cues to reduce visual workload," in *CHI '89: Proceedings of the SIGCHI conference on Human Factors in computing systems*, (New York, NY, USA), pp. 339–346, ACM Press, 1989.

[3] S. Kieffer and N. Carbonell, "Oral messages improve visual search," in *AVI '06: Proceedings of the working conference on Advanced visual interfaces*, (New York, NY, USA), pp. 369–372, ACM Press, 2006.

[4] B. Arons and E. Mynatt, "The future of speech and audio in the interface," *SIGCHI Bull.*, vol. 26, no. 4, pp. 44–48, 1994.

[5] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human-Computer Interaction*, vol. 2, no. 2, pp. 167–177, 1986.

[6] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[7] W. W. Gaver, "What in the world do we hear? An ecological approach to auditory source perception," *Ecological Psychology*, vol. 5, no. 1, 1993.

[8] B. Gygi, G. R. Kidd, and C. S. Watson, "Similarity and categorization of environmental sounds," *Perception and Psychophysics*, vol. 69, no. 6, pp. 839–855, 2007.

[9] W. W. Gaver, "The SonicFinder: An interface that uses auditory icons," *Human-Computer Interaction*, vol. 4, no. 1, pp. 67–94, 1989.

[10] M. Fernström and E. Brazil, "Human-computer interaction design based on interactive sonification hearing actions or instruments/agents," in *Proceedings of the 2004 International Workshop on Interactive Sonification*, (Bielefeld University, Germany), January 2004.

[11] M. Fernström, E. Brazil, and L. Bannon, "HCI design and interactive sonification for fingers and ears," *IEEE MultiMedia*, vol. 12, no. 2, pp. 36–44, 2005.

[12] S. Brewster, "Using non-speech sound to overcome information overload," *Displays*, vol. 17, no. 3, pp. 179–189, 1997.

[13] D. Talsma, T. J. Doty, and M. G. Woldorff, "Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration?," *Cerebral Cortex*, vol. 17, no. 3, pp. 679–690, 2007.

[14] S. Molholm, W. Ritter, D. C. Javitt, and J. J. Foe, "Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study," *Cerebral Cortex*, vol. 14, pp. 452–465, 2004.

[15] J. A. Johnson and R. J. Zatorre, "Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates," *Cerebral Cortex*, vol. 15, pp. 1609–1620, October 2005.

[16] J. Driver and C. J. Spence, "Spatial synergies between auditory and visual attention," in *Attention and Performance XV: Conscious and nonconcious information processing* (C. Umilta and M. Moscovitch, eds.), pp. 311–331, MIT Press: Cambridge, MA, 1994.

[17] C. Spence and J. Ranson, "Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention," *Perception and Psychophysics*, vol. 62, no. 2, pp. 410–424, 2000.

[18] W. Barfield, C. Rosenberg, and G. Levasseur, "The use of icons, earcons and commands in the design of an online hierarchical menu," *IEEE Transactions on Professional Communication*, vol. 34, no. 2, pp. 101 – 108, 1991.

[19] S. A. Brewster, V. P. Räty, and A. Kortekangas, "Earcons as a method of providing navigational cues in a menu hierarchy," in *HCI '96: Proceedings of HCI on People and Computers XI*, (London, UK), pp. 169–183, Springer-Verlag, 1996.

[20] S. Brewster, "Navigating telephone-based interfaces with earcons," in *HCI 97: Proceedings of HCI on People and Computers XII*, (London, UK), pp. 39–56, Springer-Verlag, 1997.

[21] S. A. Brewster, "Using nonspeech sounds to provide navigation cues," *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 3, pp. 224–259, 1998.

[22] D. Skantze and N. Dahlbck, "Auditory icon support for navigation in speech-only interfaces for room-based design metaphors," in *Proceedings of the 2003 International Conference on Auditory Display (ICAD '03)*, (Boston, MA), July 6-9 2003.

[23] M. Bussemakers and A. de Haan, "Using earcons and icons in categorization tasks to improve multimedia interfaces," in *Proceedings of the International Conference on Auditory Display (ICAD '98)*, (Glasgow, UK), 1998.

[24] M. Bussemakers and A. de Haan, "When it sounds like a duck and it looks like a dog... auditory icons vs. earcons in multimedia environments," in *Proceedings of the International Conference on Auditory Display (ICAD '00)* (P. R. Cook, ed.), pp. 184 – 189, 2000.

[25] M. P. Bussemakers, A. de Haan, and P. M. C. Lemmens, “The effect of auditory accessory stimuli on picture categorisation: implications for interface design,” in *Proceedings of the 8th HCI International Conference on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I*, (Mahwah, NJ, USA), pp. 436–440, Lawrence Erlbaum Associates, Inc., 1999.

[26] P. M. C. Lemmens, M. P. Bussemakers, and A. de Haan, “The effects of earcons on reaction times and error-rates in a dual task vs. a single task experiment,” in *Proceedings of International Conference on Auditory Display (ICAD ’00)* (P. R. Cook, ed.), p. 177183, 2000.

[27] P. M. C. Lemmens, M. P. Bussemakers, and A. de Haan, “Effects of auditory icons and earcons on visual categorization: The bigger picture,” in *Proceedings of International Conference on Auditory Display (ICAD ’01)*, July 29 - August 1 2001.

[28] W. W. Gaver, R. B. Smith, and T. O’Shea, “Effective sounds in complex systems: the ARKOLA simulation,” in *CHI ’91: Proceedings of the SIGCHI conference on Human Factors in computing systems*, (New York, NY, USA), pp. 85–90, ACM Press, 1991.

[29] E. D. Mynatt, “Designing with auditory icons: how well do we identify auditory cues?,” in *CHI ’94: Conference companion on Human Factors in computing systems*, (New York, NY, USA), pp. 269–270, ACM Press, 1994.

[30] P. Lucas, “An evaluation of the communicative ability of auditory icons and earcons,” in *Proceedings of the International Conference on Auditory Display (ICAD ’94)*, 1994.

[31] C. A. Sikora, L. Roberts, and L. Murray, “Musical vs. real world feedback signals,” in *CHI ’95: Conference companion on Human Factors in computing systems*, (New York, NY, USA), pp. 220–221, ACM Press, 1995.

[32] R. Graham, “Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application,” *Ergonomics*, vol. 42, p. 1233 1248, September 1999.

[33] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, “A detailed investigation into the effectiveness of earcons,” in *Proceedings of International Conference on Auditory Display (ICAD ’92)*, (Santa Fe Institute, Santa Fe), pp. 471 – 498, Addison-Wesley, 1992.

[34] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, “An evaluation of earcons for use in auditory human-computer interfaces,” in *CHI ’93: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 222–227, ACM Press, 1993.

[35] J. Edworthy, “Does sound help us to work better with machine? a commentary on Rauterberg’s paper ’about the importance of auditory alarms during the operation of a plant simulator’,” *Interacting with Computers*, vol. 10, pp. 401–409, 1998.

[36] S. Brewster, “The design of sonically-enhanced widgets,” *Interacting with Computers*, vol. 11, no. 2, pp. 211–235, 1998.

[37] E. D. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis, “Designing audio aura,” in *CHI ’98: Proceedings of the SIGCHI conference on Human Factors in computing systems*, (New York, NY, USA), pp. 566–573, ACM Press/Addison-Wesley Publishing Co., 1998.

[38] S. A. Brewster, G. Leplatre, and M. G. Crease, “Using non-speech sounds in mobile computing devices,” in *Proc. of the First Workshop on Human Computer Interaction with Mobile Devices* (J. C., ed.), (Department of Computing Science, University of Glasgow, Glasgow, UK), pp. 26 – 29, 1998.

[39] G. Leplatre and S. Brewster, “Designing non-speech sounds to support navigation in mobile phone menus,” in *Proceedings of the 6th International Conference on Auditory Display (ICAD ’00)* (P. R. Cook, ed.), pp. 190 – 199, 2-5 April 2000.

[40] S. Holland, D. R. Morse, and H. Gedenryd, “AudioGPS: Spatial audio navigation with a minimal attention interface,” *Personal Ubiquitous Computing*, vol. 6, no. 4, pp. 253–259, 2002.

[41] N. Sawhney and C. Schmandt, “Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments,” *ACM Trans. Computer-Human Interactions*, vol. 7, no. 3, pp. 353–383, 2000.

[42] E. Murphy, *Designing Auditory Cues for a Multimodal Web Interface: A Semiotic Approach*. PhD thesis, Queen’s University, Belfast, Ireland, 2007.

[43] E. Mynatt, “Transforming graphical interfaces into auditory interfaces for blind users,” *Human-Computer Interaction*, vol. 12, pp. 7–45., 1997.

[44] E. D. Mynatt and G. Weber, “Nonvisual presentation of graphical user interfaces: contrasting two approaches,” in *CHI ’94: Proceedings of the SIGCHI conference on Human Factors in computing systems*, (New York, NY, USA), pp. 166–172, ACM Press, 1994.

[45] S. Morley, H. Petrie, A. O’Neill, and P. McNally, “Auditory navigation in hyperspace: design and evaluation of a non-visual hypermedia system for blind users,” *Behaviour and Information Technology*, vol. 18, pp. 18–26, January 1999.

[46] S. Goose and C. Möller, “A 3D audio only interactive web browser: using spatialization to convey hypermedia document structure,” in *MULTIMEDIA ’99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, (New York, NY, USA), pp. 363–371, ACM Press, 1999.

[47] E. Murphy, R. Kuber, P. Strain, G. McAllister, and W. Yu, “Developing sounds for a multimodal interface: conveying spatial information to visually impaired web users,” in *Proceedings of the 13th International Conference on Auditory Display (ICAD ’07)*, pp. 348–355, June 26-29 2007.

[48] M. Slater and M. Usoh, “Presence in immersive virtual environments,” in *Virtual Reality Annual International Symposium*, pp. 90–96, 1993.

[49] M. Grohn, T. Lokki, and T. Takala, “Comparison of auditory, visual, and audiovisual navigation in a 3D space,” in *Proceedings of the 9th International Conference on Auditory Display (ICAD ’03)*, 2003.

[50] Y. Zhang, T. Fernando, H. Xiao, and A. R. L. Travis, “Evaluation of auditory and visual feedback on task performance in a virtual assembly environment,” *Presence*, vol. 15, no. 6, pp. 613–626, 2006.

[51] G. W. Edwards, W. Barfield, and M. A. Nussbaum, “The use of force feedback and auditory cues for performance of an assembly task in an immersive virtual environment,” *Virtual Reality*, vol. 7, pp. 112–119, 2004.